**Name:** Rushiraj Suwarnkar  
**Roll No:** 281018  
**Batch:** A1

**Assignment 4**

**Statement:**

In this assignment, we perform **visualization of data** using various types of plots to gain insights into datasets used in Assignment 1 (admission.csv) and Assignment 2 (heart.csv). We use the following chart types to represent the data:

* Scatter Plot
* Bar Plot
* Box Plot
* Pie Chart
* Line Chart

**Objective:**

1. To visually explore and analyze datasets using Python (Matplotlib & Seaborn).
2. To understand the relationship between variables.
3. To identify trends, distributions, and outliers using graphical methods.

**Resources Used:**

* **Software:** Jupyter Notebook
* **Language:** Python
* **Libraries:** pandas, matplotlib, seaborn

**Dataset 1: admission.csv**

This dataset contains student information for graduate school admissions prediction, with features like GRE, TOEFL, University Rating, SOP, LOR, CGPA, Research, and Chance of Admit.

**Operations Performed:**

* **Scatter Plot:**
  + Plot between GRE Score and Chance of Admit
  + Shows positive correlation
* **Bar Plot:**
  + Count of applicants with and without Research
  + Helps compare research vs. non-research applicants
* **Box Plot:**
  + Box plot of CGPA
  + Identifies outliers in CGPA scores
* **Pie Chart:**
  + Distribution of University Rating
  + Visualizes student proportion in each rating
* **Line Chart:**
  + Line chart for GRE Score vs. TOEFL Score
  + Displays trend in language vs aptitude scores

**Dataset 2: heart.csv**

This dataset contains patient records with various features like Age, Chest Pain, Cholesterol, Heart Rate, and heart disease status.

**Operations Performed:**

* **Scatter Plot:**
  + Age vs. Chol (Cholesterol)
  + Helps detect pattern between age and cholesterol level
* **Bar Plot:**
  + Count of each ChestPain type
  + Compares frequency of different types of chest pain
* **Box Plot:**
  + Box plot for MaxHR (Maximum Heart Rate)
  + Identifies heart rate distribution and outliers
* **Pie Chart:**
  + Distribution of AHD (Heart Disease: Yes/No)
  + Shows proportion of people with and without heart disease
* **Line Chart:**
  + Line chart showing Age vs. MaxHR
  + Useful to observe age-based change in heart rate

**Advantages:**

1. Easy to interpret and explain data through visual patterns.
2. Outliers and trends are quickly visible.
3. Helps in feature selection for ML models.

**Disadvantages:**

1. Graphs may become unclear with too many categories.
2. Pie charts are less effective with small value differences.
3. Interpretation requires domain knowledge in medical datasets.

**Conclusion:**

This assignment helped in understanding how different plots can reveal patterns, anomalies, and insights from the data. Using real-world datasets like heart.csv and admission.csv, we practiced scatter plots, bar plots, box plots, pie charts, and line charts effectively. These visualization tools are critical in the data analysis and preprocessing stages of any data science or machine learning project.